Data analysts ask increasingly sophisticated questions of larger and more complex datasets, statistical models are becoming both more complex and more time-consuming to estimate. For problems as disparate as the production of astronomical catalogues, analysis of internet user data, to the meta-analysis of randomized controlled trials in development economics.

As models grow in complexity, the need to interrogate their assumptions, to propagate uncertainty amongst their components, and to perform non-parametric checks on their data sensitivity grows commensurately, but so does the computational cost of doing so.

Classical procedures such as Markov Chain Monte Carlo (MCMC) or the bootstrap, which require evaluating a model for many distinct parameter values or datasets, respectively, can be prohibitively expensive.

My research focuses on applications of sensitivity analysis that use only properties of a single model fit to extrapolate to alternatives without expensive re-running. My work is conceptually unified around a single theme, but disparate in applications.

My guiding motivation is uncertainty quantification, providing easy-to-use, general tools that allow analysts to reason about both aleatoric and epistemic uncertainty in their analysis tasks.

Sensitivity analysis is traditionally construed both more broadly and more narrowly than the methods I investigate. Traditional sensitivity analysis encompasses both local methods, which are based on extrapolations from a given model fit, and global methods, which typically require fitting a model multiple times at a variety of configurations (though exceptions exist in cases with special structure). Traditionally, sensitivity analysis focuses only on assessing whether small changes in arbitrary modeling decisions can affect the outcome. Due to the computational expense of re-fitting many modern statistical models, and the relative ease of computing derivatives (especially with modern automatic differentiation tools), I focus on local sensitivity analysis. And a core theme of my work is that the domain of sensitivity analysis in fact encompasses much of classical frequentist uncertainty quantification, and can provide valuable insights into Bayesian uncertainty quantification as well.

Prior Sensitivity in Bayesian Analysis

Bayesian analysis allows analysts to reason coherently about unknown parameters, but only if the user specifies a complete generating process for the parameters and data, including both prior distributions for the parameters and precise likelihoods for the data. Often aspects of this model are at best a considered simplifcation, and at worst chosen only for computational convenience. It is critical to ask whether the analysis would have changed substantively had different modeling choices been made.

Bayesian Nonparametrics

A commonly asked question in unsupervised clustering is how many distinct clusters are present in a dataset. Discrete Bayesian nonparametrics allows the question to be addressed using Bayesian inference, but one must specify a prior on how distinct clusters are generated. A particularly common choice is the stick-breaking representation of a Dirichlet process prior, a mathematical abstraction arguably better justified by its mathematical convenience than its realism. The prior must be specified in terms of random stick lengths to be broken off successively, the lengths of the sticks determining the a priori cluster sizes. The standard approach is to model the stick lengths with $Beta(1, \alpha)$ distribution, the α being a scalar tuning parameters. Other classes of distributions are possible in principle but hardly ever considered in practice, in part because the Beta distribution enjoys some computational conveniences.

In CITE, we provide sensitivity measures that allow the user to explore alternative stick breaking distributions from a single fit using the standard and convenient Beta prior. We linearly approximating the dependence of the optimum on the functional form of a parameterized class of priors. A natural parameterized class is the set of $Beta(1, \alpha)$ distributions (parameterized by α), but we also consider arbitrarily functional perturbations. In current work in progress, we evaluate the worst-case perturbations. On a real-world clustering problem, a human genome dataset, we find that the number of distinct inferred populations is in fact quite sensitive to the prior.

Partial Pooling in Meta-analysis

A popular form of meta-analysis is to place a hierarchial model on a set of related experimental results, which both "shrinks" the individual estimates towards a common mean, potentially decreasing mean squared error, and allowing direct estimation of the average effect and diversity of effects. These advantages come at the cost of positing a precise generative process for the effects in question, however, and it is reasonable to interrogate whether the estimation procedure is robust to varaibility in these effects. In CITE, we apply sensitivity analysis to a published meta-analysis of the effectiveness of microcredit interventions in seven developing countries. We find that the conclusion are highly sensitive to the assumed covaraince structure between the base level of business profitability and the microcredit effect, a covariance which is a priori difficult to ascertain. In this way, we were able to easily diagnose a conceptual problem in a model which was time-consuming to fit.

Hyperparameter Sensitivity for MCMC

A classical result in Bayesian sensitivity analysis states that derivatives of posterior expectations take the form of particular posterior covariances. The resulting sensitivities can be automatically computed in a black-box manner when the posterior is implemented in software that supports automatic differentiation, such as the popular Hamiltonian Monte Carlo sampler and modleing language,

Stan. I have written an R package CITE that allows Stan users to specify a "hyperparameters" modeling block, from which one can automatically compute hyperparameter sensitivity from a single MCMC run with no additional computation. I apply these principles in a related work on frequentist variance below.

Data Sensitivity, Cross Validation, and Frequentist Variance

Frequentist variability is ultimately concerned with the value of an estimation procedure if the data were different than that observed. A classical mainfestiation of this idea is the nonparametric bootstrap, which estimates frequentist variability by evaluating a particular estimator at pseudo-datasets with observations draw with replacement from the observed dataset. Similarly, cross-validation (CV) in its various forms evaluates how a statistical procedure performs on data that were not included as part of estimation, and can be thought of as a non-parametric estimator of the bias induced by evaluating a loss function using the same data that were used to fit a model.

Both the bootstrap and CV require re-fitting a model with new, nearby datasets multiple times. When the model is differentiable, and model re-fitting is expensive, it can be advantageous to approximate the effect of re-fitting rather than perform actual re-fitting. One way of doing so it so is to perform a Taylor series expansion of the estimator, as a funciton of the empirical distribution, around the original empirical distribution. This is the core concept behind the related classical tools known as the "infinitesimal jackknife," "von-Mises Expansion," and "empirical influence function," though until recently these differnetial approximations were used most prominently to facilitate theoretical analysis.

We, and several other authors, obsevered that these differnetial methods could speed up the evaluation of cross validation in large machine learning models which are expensive to re-fit. In CITE, we bridged the gap from some of the classical literature, providing finite-sample accuracy bounds for approximate leave-k-out cross validation, even when the derivatives of the objective function are unbounded.

A follow-on work in progress (CITE) expands the results to higher-order expansions and to larger perturbations, including k-fold CV and the bootstrap. They key to all this work is a set complexity condition, in light of which it is clear that one can provide accuracy bounds uniformly over small perturbations, and over randomly-chosen large perturbations, but not for uniform bounds over large perturbations. Because we show that the linear approximation approaches the bootstrap closer than the bootstrap approaches the truth, our work should allow for practical differential approximations to prohibitively expensive procedures such as the bootstrap-after-the-bootstrap.

Propagation of Uncertainty in Variational Bayes